Deploying Splunk on Amazon Web Services

Simeon Yep
Strategic Alliances
Nate Kwong
Senior SE
Bill Bartlett
Senior SE

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Objective:

Integrate your Splunk Enterprise deployment with Amazon Web Services (AWS)

#1

MQ Leader

Presenters

Bill Bartlett

- Senior SE, AWS
- Seattle

Nate Kwong

- Senior SE, Majors
- SF

Simeon Yep

- Director, Alliances
- SF

Agenda

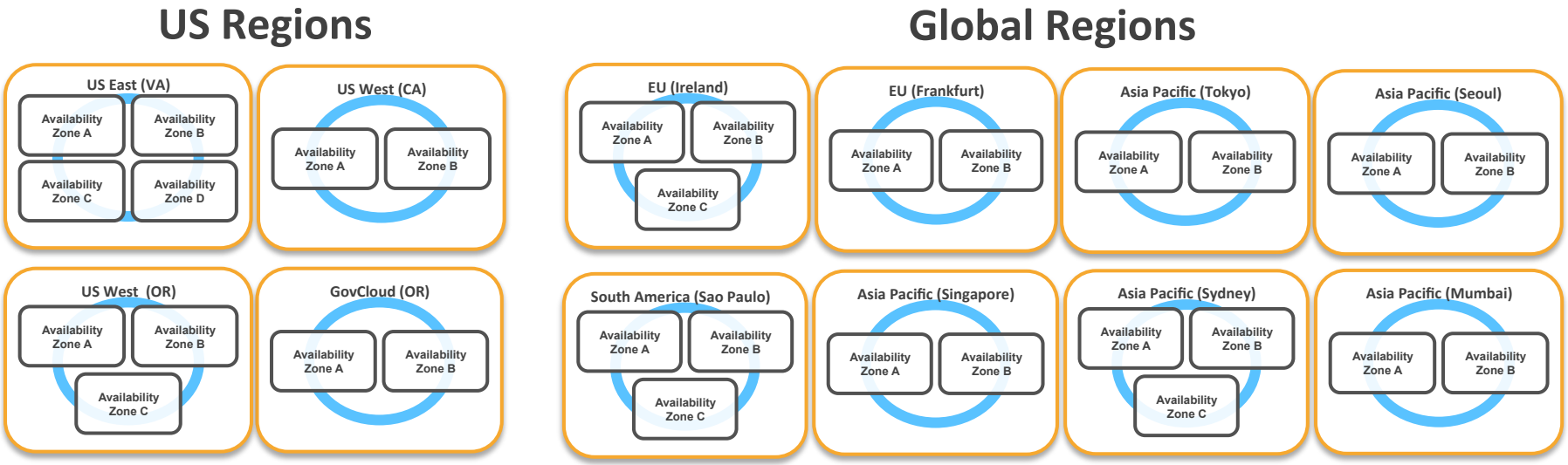
- Infrastructure: AWS Elastic Compute Cloud (EC2)
- Deployment Examples & leveraging AWS features
- AWS Provisioning and Automation
- Splunk App for AWS Demo

AWS EC2 Infrastructure

Global Infrastructure

Global Infrastructure

AWS Regions & Availability Zones

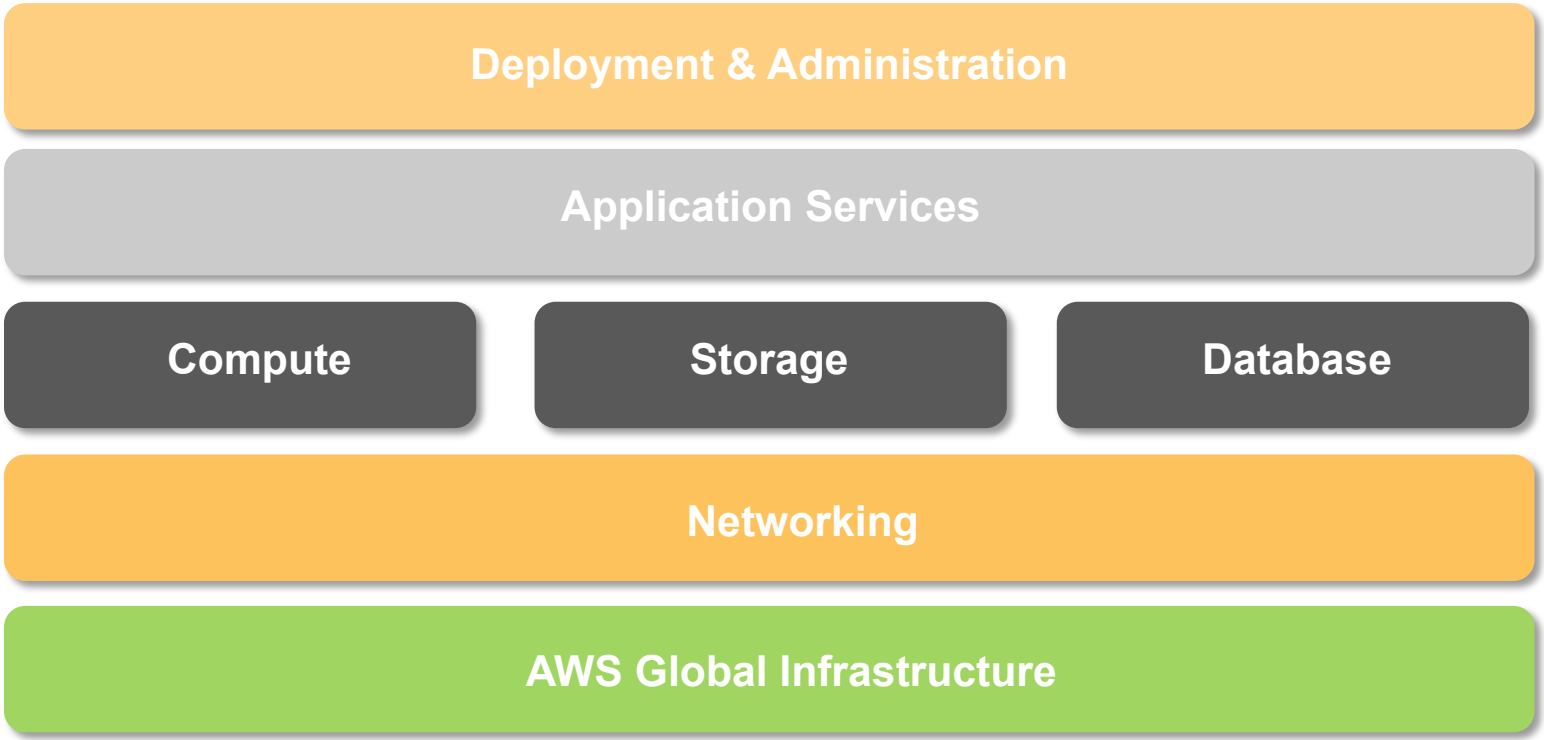


AWS China (Beijing) Region – 2 Availability Zones https://www.amazonaws.cn/

Customer Decides Where Applications and Data Reside

Note: Conceptual drawing only. The number of Availability Zones may vary.

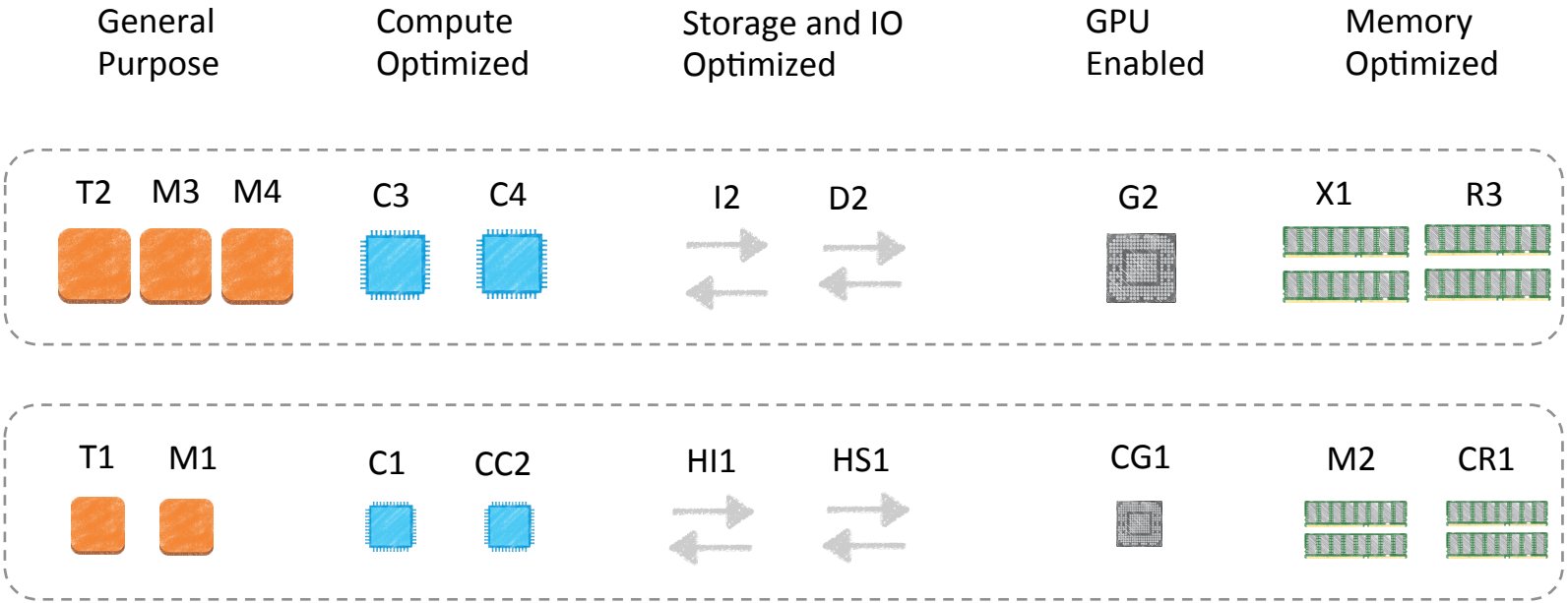
Broad and Deep Services to Support any Cloud Workload



Amazon Web Services EC2

- Amazon Elastic Compute Cloud (EC2
- Pay-as-you-go pricing model
- Splunk is easily deployed in Amazon

Broad Set of Compute Instance Types…



Typical User Scenario

1. Sign-up for an AWS account (use AWS IAM – Identity and Access Management)
2. Launch an EC2 instance (via user chosen tool such as GUI, CLI, or external)
3. Use key credentials to access the EC2 instance
4. Install Software/Splunk

Splunk + AWS + CloudFormation
Ready in minutes…

Splunk and Hardware

- Splunk consumes high I/O due to indexing and searching
- Load != GB/day
- Search drives a large portion of the load
  - Rare vs. Sparse vs. Reporting
  - Real-time vs. Historic
- Rule of thumb – up to 300 GB/day
  - Reference servers can index much more, up to TB/day with no search load
- Virtualized systems incur some overhead, but work well if tuned correctly

Instances

- Instance type
  - Pricing: Spot vs. On-demand vs. **Reserved**
  - Family: **Storage** vs. **Compute** vs. GPU vs. Memory vs. **General Purpose**
  - Generation: **Current** vs. Previous
- Instance size
  - Workload size: compute units, memory, storage
  - Micro, Small, Medium, Large, Extra Large (XL)
    - Multiple XL sizes: xlarge, 2xlarge, 4xlarge, 8xlarge
  - 4XL general purpose provides similar performance to a reference server
    - 50-300GB/day indexing and searching

Instance Storage

- Instances have ephemeral storage (Current Gen has SSDs)
  - General Purpose instances have GBs to TBs
  - Storage Optimized instances have up to 48 TB!
  - Data is lost when the instance dies
- EBS – Elastic Block Storage
  - Persistent block level storage volumes for use with EC2 instances
  - Cost associated – 1 TB costs $100/month
  - Data is not lost when instance dies – can be remounted with new instance
  - For storage needs larger than 16 TB, RAID required
  - Built-in resiliency – data is backed up
- S3 – Simple Storage Service
  - Online cloud storage service (files, data, etc…)
  - Need this for backup purposes (Snapshots)
  - Can also be used as a data feed for Splunk, TA available

Storage Best Practices

- Single instances or non-replicated distributed deployments:
  - Use EBS volumes for indexes and the OS/software
  - RAID can be an extra measure of reliability, but will consume CPU
  - Use snapshots to backup the instance (S3)
  - IOPS optimized provides benefits
  - XFS preferred (customer feedback)
  - c4 (compute optimized) Instances will require storage

Instance Selection

- How can I make my deployment resilient?
  - Option 1: EBS
  - Option 2: Index Replication
  - Option 3: Data Cloning (Index and Forward)
- Instance selection should factor in resiliency, use-case, and cost
- Index Replication (IR)
  - Replication requires more instances as data is stored twice
  - Does not require EBS for indexes
  - Major driver is instance cost as you leverage ephemeral storage

Instance Selection Exercise

- 1 TB/day Distributed Deployment
  - EBS backed storage for availability
  - No replication
- AWS Calculator spreadsheet available

1000 GB/day + EBS setup			
Instance Type	Total Instances		Total Cost (Monthly)
c4.4xlarge	4	192	$ 2,410.82
c4.4xlarge	1		$ 602.70
	Storage Cost	Retention (Days)	
EBS - (1 TB/day, 50% compression)	$0.1 GB/Month	192	$ 9,600.00
			$ 12,613.52

Instance Selection Exercise

- Retention values for EBS backed deployments significantly drive cost

1000 GB/day + EBS setup			
Instance Type	Total Instances		Total Cost (Monthly)
c4.4xlarge	4	192	$ 2,410.82
c4.4xlarge	1		$ 602.70
	Storage Cost	Retention (Days)	
EBS - (1 TB/day, 50% compression)	$0.1 GB/Month	192	$ 9,600.00
			$ 12,613.52

1000 GB/day + EBS setup			
Instance Type	Total Instances	Retention (Days)	Total Cost (Monthly)
c4.4xlarge	4	15	$ 2,410.82
c4.4xlarge	1		$ 602.70
	Storage Cost	Retention (Days)	
EBS - (1 TB/day, 50% compression)	$0.1 GB/Month	15	$ 750.00
			$ 3,763.52

Instance Selection Exercise

- 1 TB/day Distributed Deployment
  - Index Replication enabled (Double the indexers and add 1 Administrative node)
- Index Replication offers immediate search capability with SF/RF
- Differences:
  - $5k
  - Increased availability, higher performance

1000 GB/day (IR + SHC)	1 TB/day + Index Replication + Search Clustering + CM/Deployer		
Instance Type	Total Instances	Retention (Days)	Total Cost (Monthly)
d2.4xlarge	8	192	$ 14,837.76
c4.4xlarge	4		$ 2,410.82
			$ 17,248.58

Instance Selection

Distributed Deployments

Using Index Replication (IR)

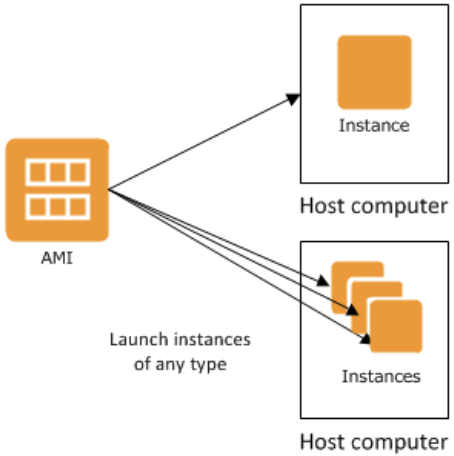
- Local ephemeral storage (SSDs) may perform better than EBS
- Search/Replication Factor determines availability of data for searching
- IR adds load and requires more servers and storage

Using EBS volumes, no IR

- Typically fewer instances to manage vs. IR
- Search Availability is driven by the capability to remount a volume to a new instance (automatically or manually)
- Cost can be largely driven by retention and daily volume

Amazon Machine Image (AMI)

- Amazon Machine Image (AMI) preferences for Splunk
  - Amazon Linux based
  - Best Performance
  - Cost Effective (extra $$ for Windows)
- AMIs available for download
  - Splunk Enterprise
  - Hunk
    - Hunk + EMR baked into Marketplace



Best Practices

- Custom AMI creation
  - Create your own AMI using Linux based or Splunk provided
  - Leverage current configuration tooling with AMI (don't have to use deployment server, but can be very helpful)
- Authentication and Authorization
  - Policies will dictate what you can or cannot use
  - LDAP/AD will require an SSL tunnel
  - Other options: scripted input or proxying (SSO)
  - SAML (Okta, AD FS, Ping, Azure AD)
  - NOTE – SSO methods still require role information
- Security
  - SSL everywhere + private network
  - Install your own certificates

Best Practices

- Search Head Clustering
- Deploy to the same AWS Region
  - Replication and searches across Regions can be a challenge
- Monitor from outside of the Region/AZ
  - Offers additional resiliency
- Use a Virtual Private Cloud (VPC)

General Guidelines

Follow Best Practices for Architecting and Sizing: **Load=Searching+Indexing**

Indexers (50-300GB/day)

- c4.4xlarge 16 vCPU, 30 GB RAM
- d2.4xlarge 16 vCPU, 122 GB RAM
- c4.8xlarge 36 vCPU, 60 GB RAM

*These are all starting points! Splunk can index and search more OR less depending on overall load.

Search Heads (8+ users)

- c4.4xlarge 16 vCPU, 30 GB RAM
- c4.8xlarge 36 vCPU, 60 GB RAM

Cluster Master or Deployment Server

- c4.xlarge 4 vCPU, 7.5 GB RAM
- c4.2xlarge 8 vCPU, 15 GB RAM

License Master

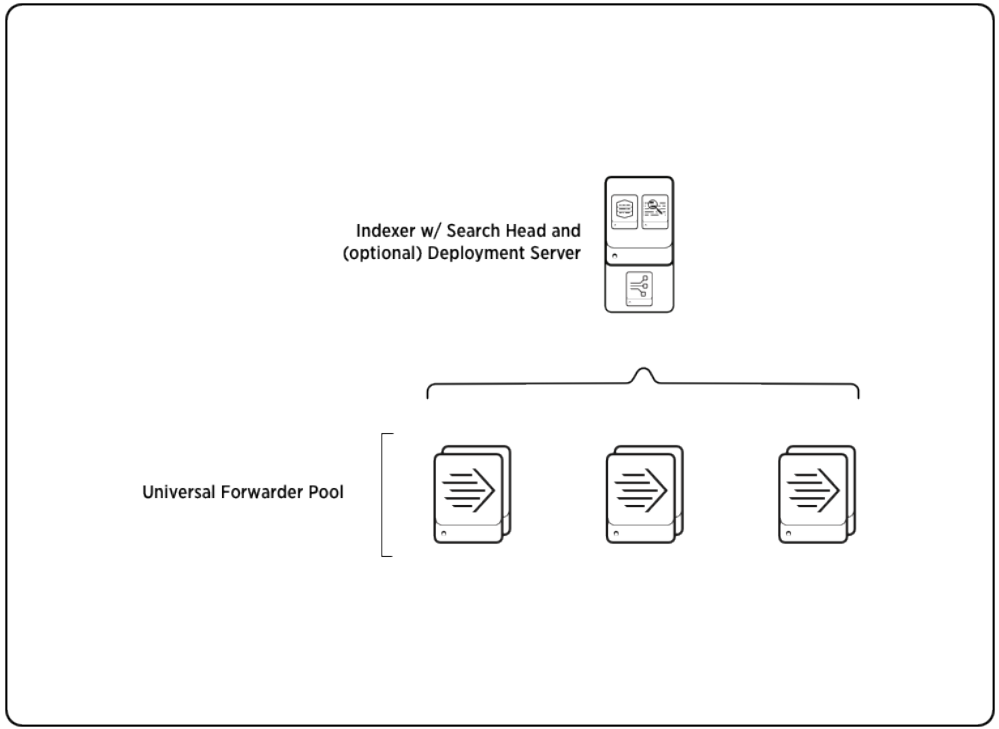
- c4.large 2 vCPU, 3.75 GB RAM
- c4.xlarge 4 vCPU, 7.5 GB RAM

Architecture & Deployment Examples

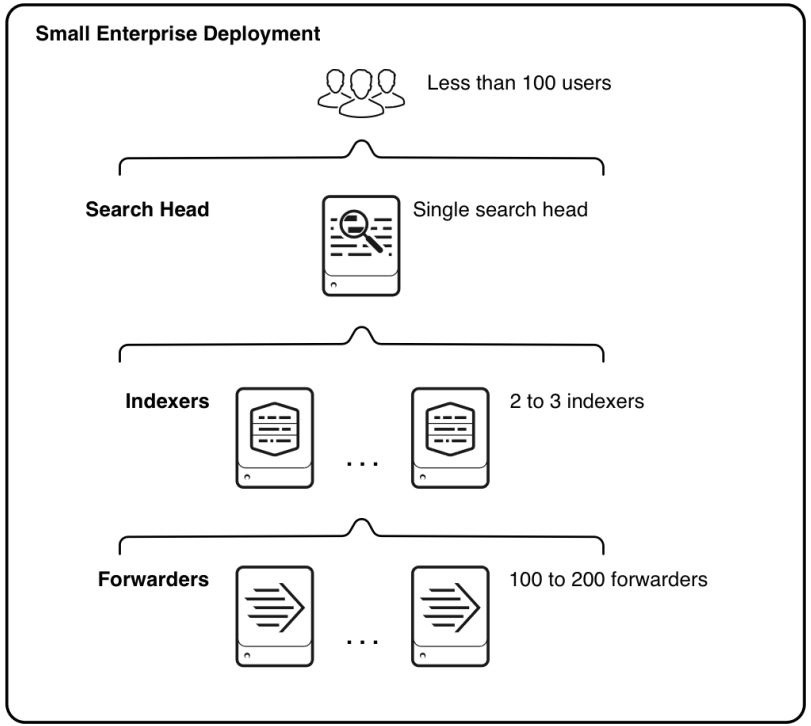
Architecture Examples

- Single
- Distributed
- Distributed with Index Replication

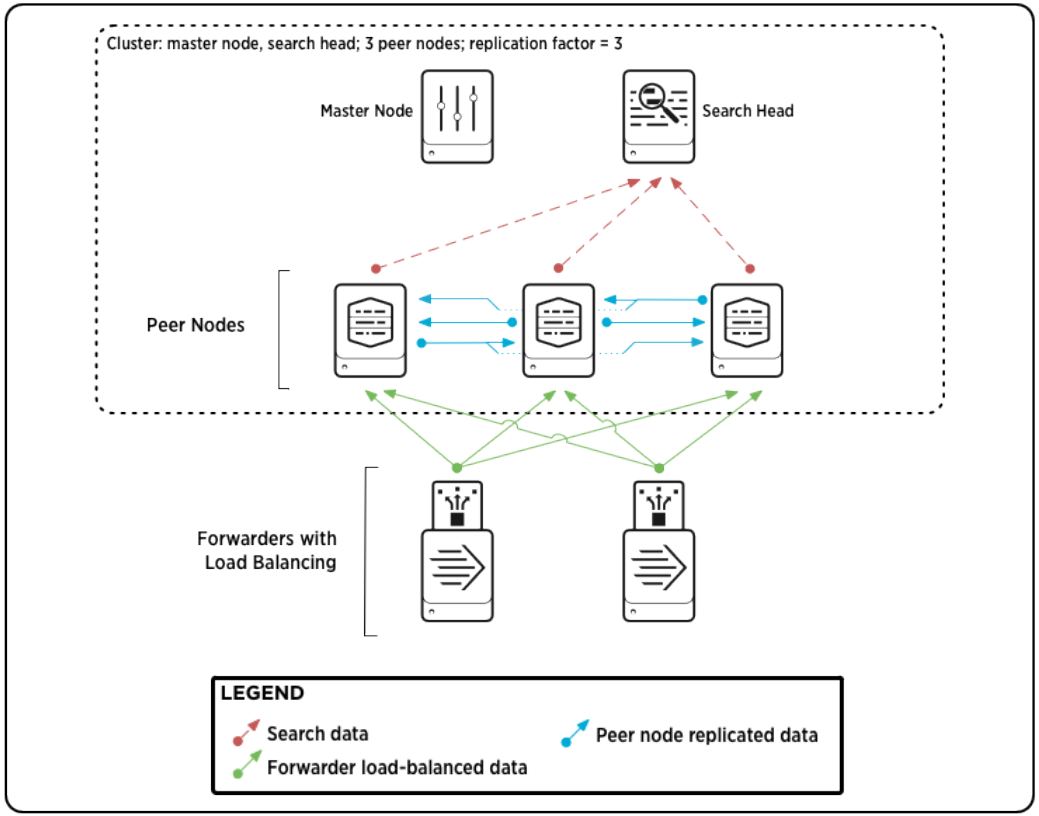
Single Server



Distributed



Distributed with Index Replication



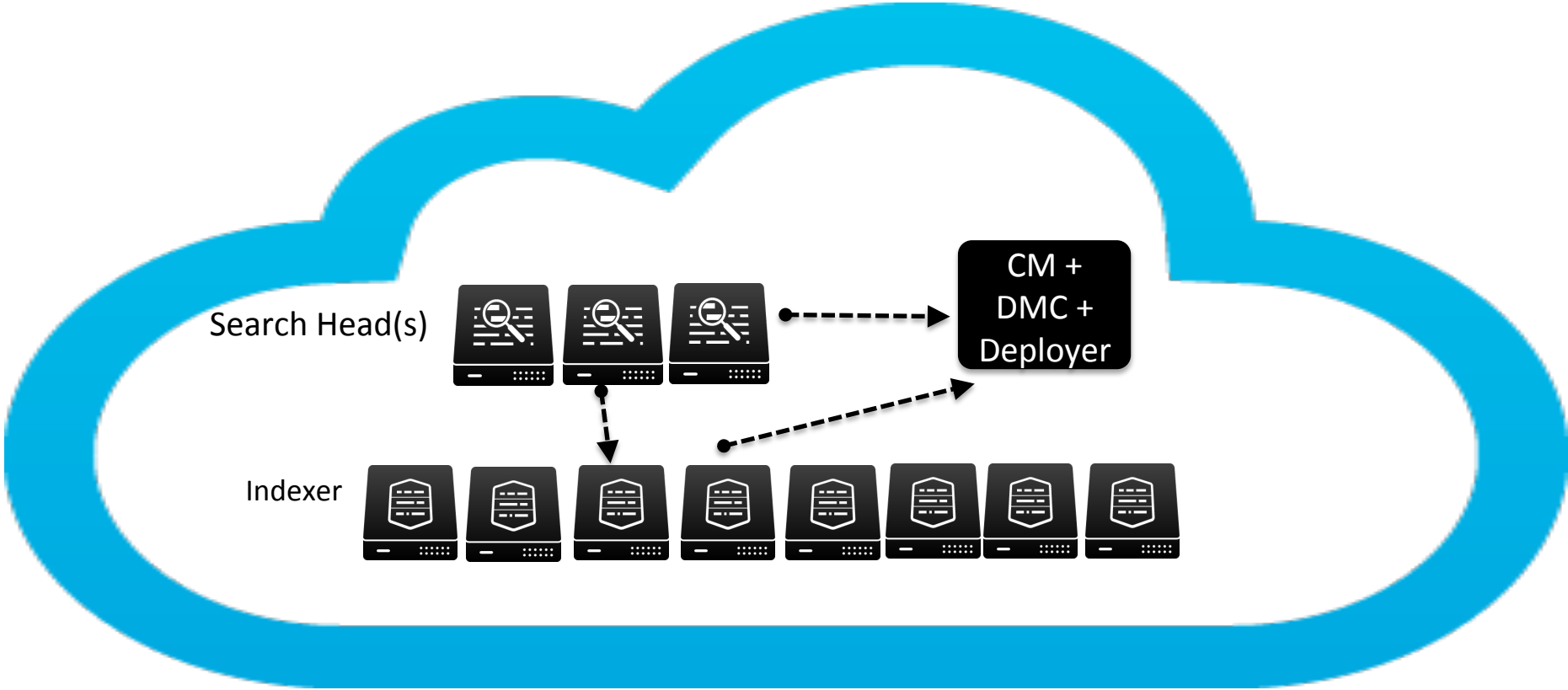
Single Server

- Use case: Searching, Reporting and Analytics
- Up to 300+ GB/day indexing with common search loads
  - For heavy reporting and analytics, decrease indexing volume
- C4.8xlarge instance
  - EBS volumes configured to support retention needs
- Up to 16 concurrent users

Distributed Deployment

- Use Case: Application Management, Security Forensics
- Up to 1 TB/day indexing with common search loads
- Distributed deployment with Index Replication (2 SF, 3 RF)
- 8 - d2.4xl instances with 24 TB ephemeral storage (indexers)
- 3 - c3.4xlarge instance (search cluster)

Deployment



Example Architectures

Use case and requirements influence final setup, but there is no right or wrong way

Using EBS (elastic block store)

- 20 GB/day
  - c4.2xlarge (single instance)
- 100 GB/day
  - c4.4xlarge (single instance)
- 300 GB/day
  - c4.4xlarge
  - c4.8xlarge
- 500 GB/day
  - c4.4xlarge as indexer (3)
  - c4.4xlarge as search head (1)
- 1000 GB/day
  - c4.4xlarge as indexer (6)
  - c4.8xlarge as search head (1)
- 1500 GB/day
  - c4.4xlarge as indexer (9)
  - c4.8xlarge as search head (1)

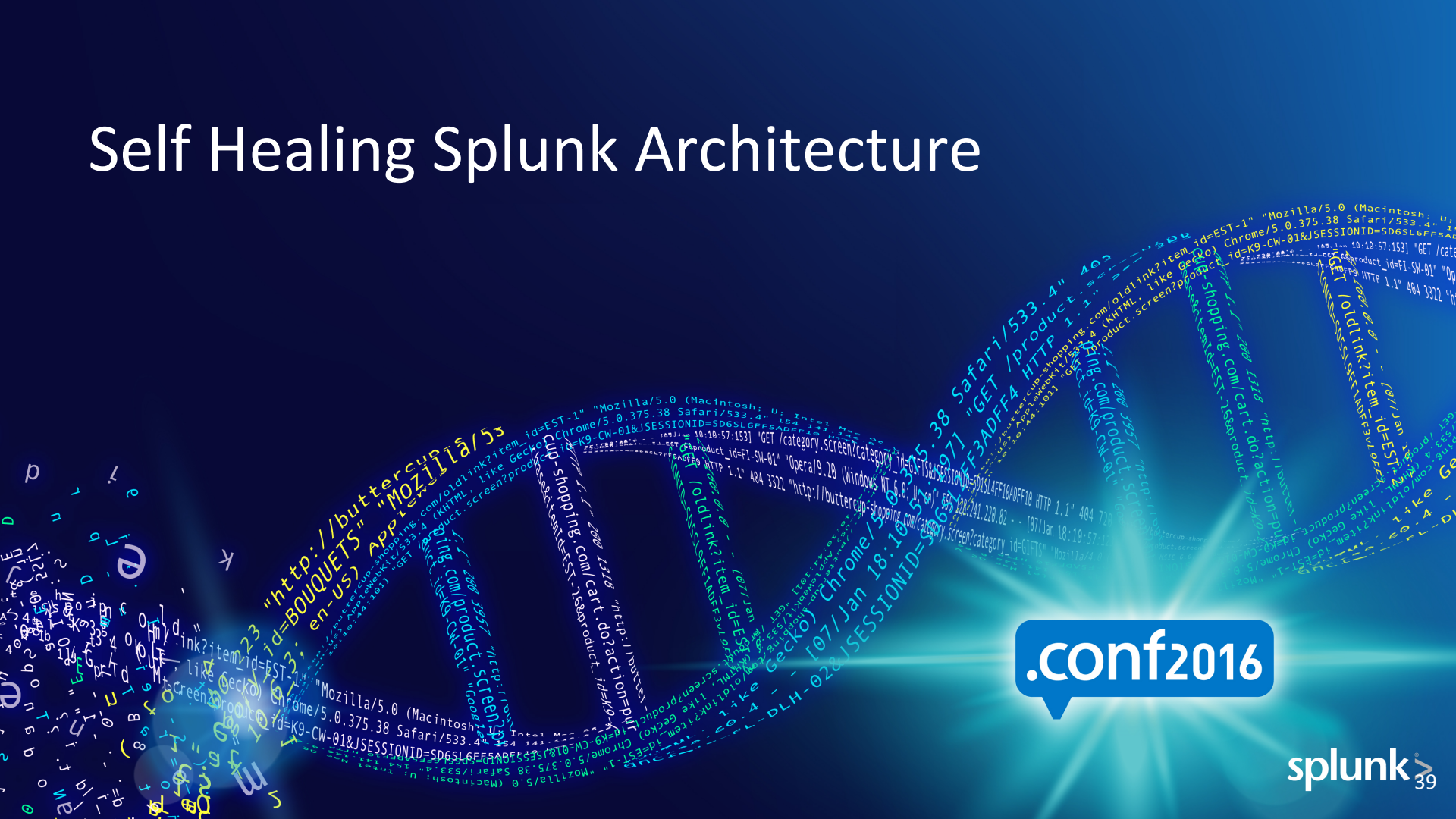
Example Architectures

Use case and requirements influence final setup, but there is no right or wrong way

Using Index Replication

- 100 GB/day
  - d2.2xlarge as indexer (2)
  - c4.2xlarge as search head (1)
  - c4.xlarge as CM
- 500 GB/day
  - d2.4xlarge as indexer (3)
  - c4.4xlarge as search head (1)
  - c4.xlarge as CM
- 1000 GB/day
  - d2.4xlarge as indexer (6)
  - c4.8xlarge as search head (1)
  - c4.2xlarge as CM (1)
- 1500 GB/day
  - d2.4xlarge as indexer (9)
  - c4.8xlarge as search head (1)
  - c4.2xlarge as CM (1)

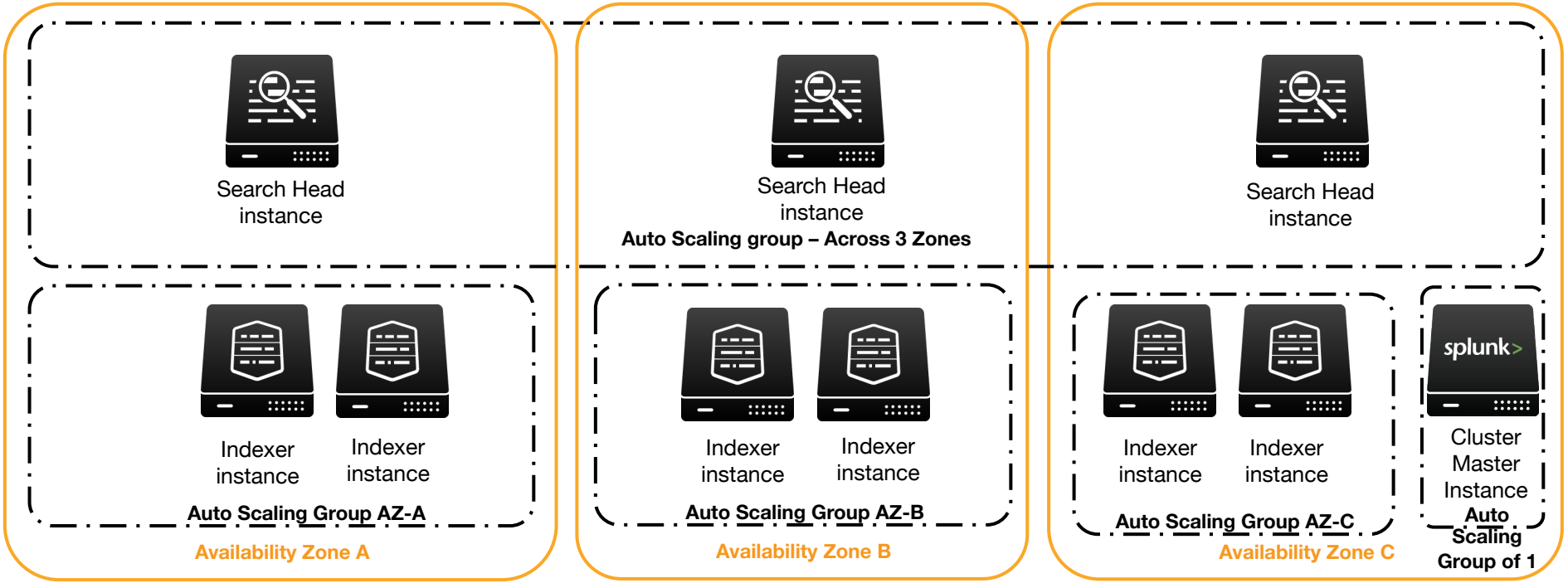
Self Healing Splunk Architecture



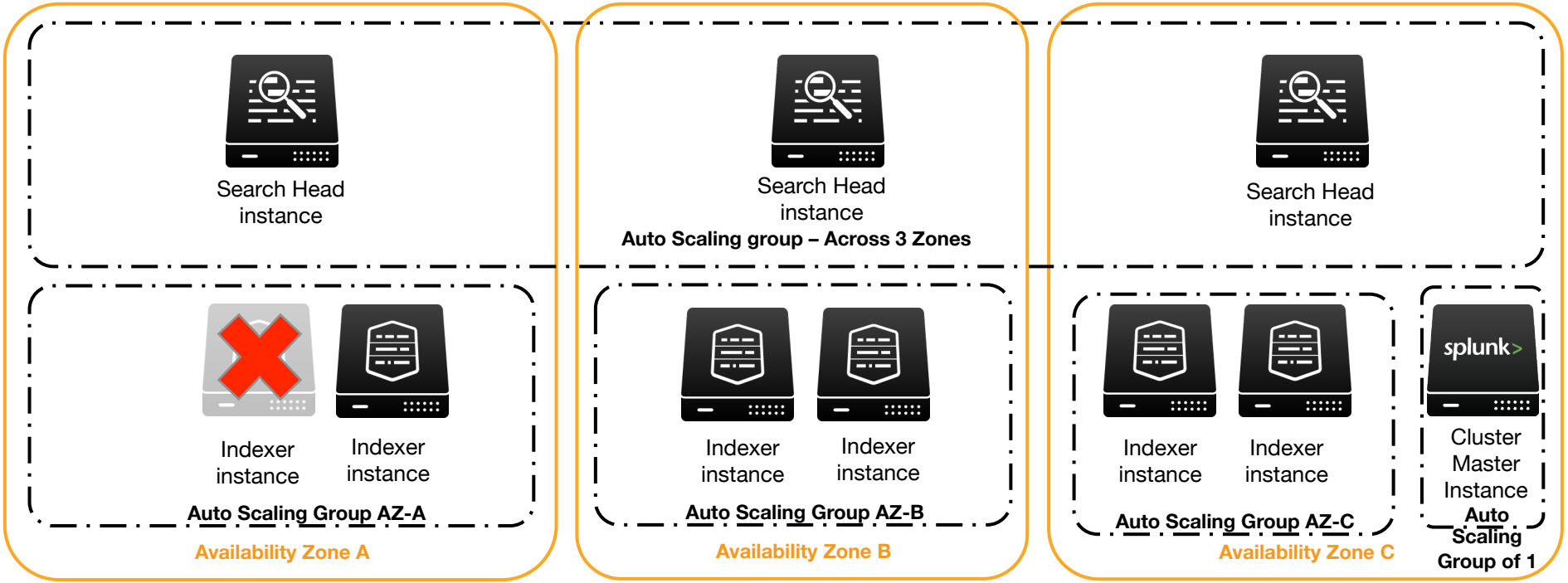
AWS Auto Scaling

- Automatically replace unhealthy EC2 instances
- Multiple Auto Scaling Policies
  - Maintain a fixed number of EC2 Instances (*recommended for Splunk Indexers*)
  - Performance metrics
  - Time based
  - Manual Scaling

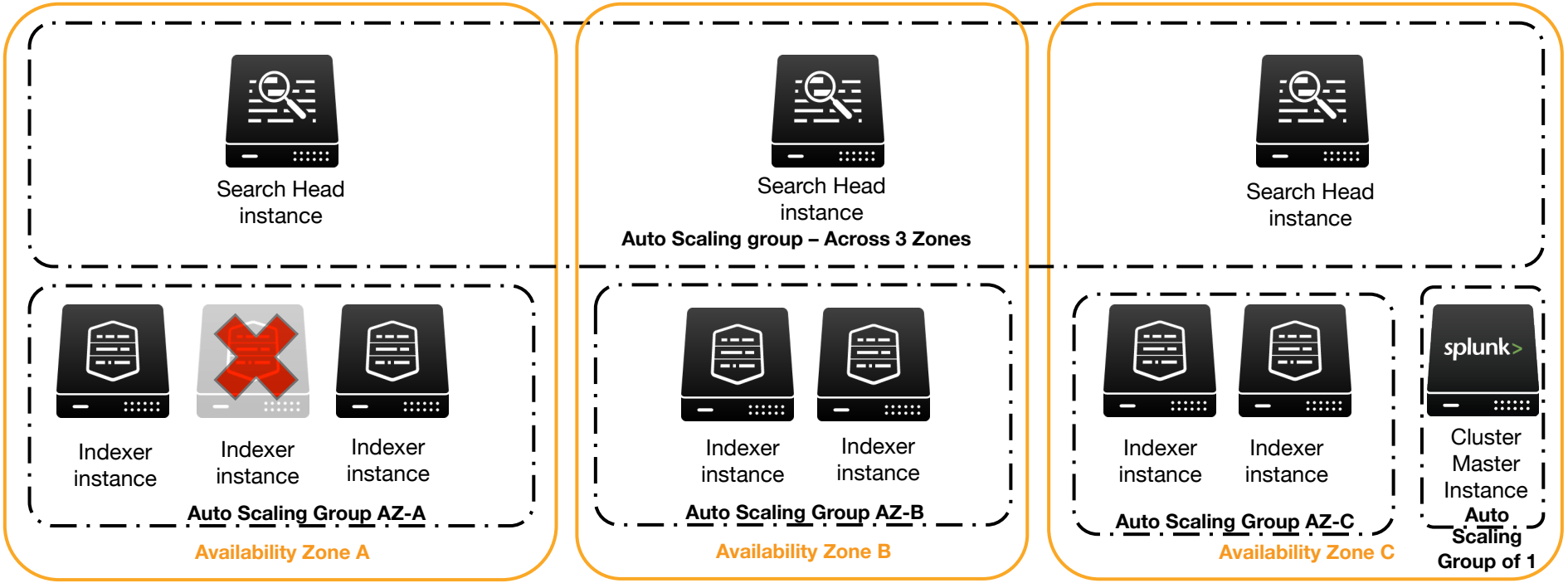
Architecture Diagram (Splunk + AWS)



Architecture Diagram (Splunk + AWS)



Architecture Diagram (Splunk + AWS)



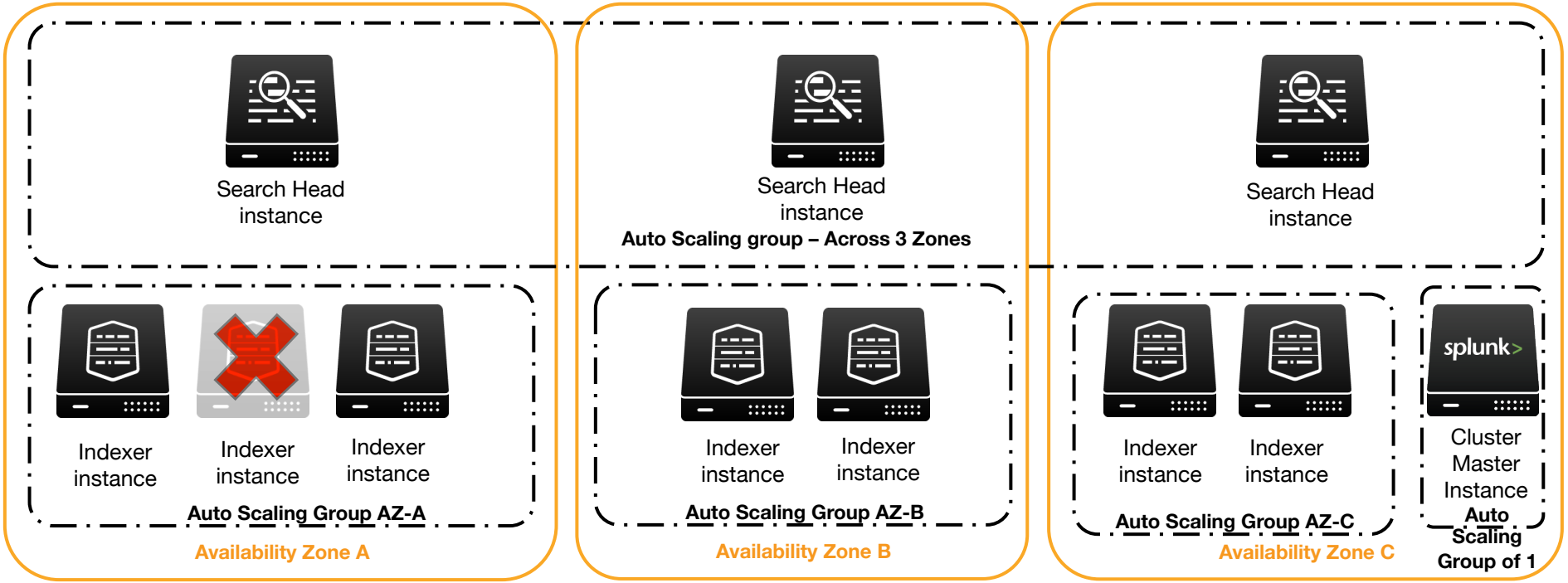
Splunk Indexer Clustering with Auto Scaling

- Multisite clustering
  - Replicate a copy of your data to multiple sites
  - *Hint: AWS Availability Zone = Splunk Site*
- Separate Auto Scaling Groups for each Availability Zone

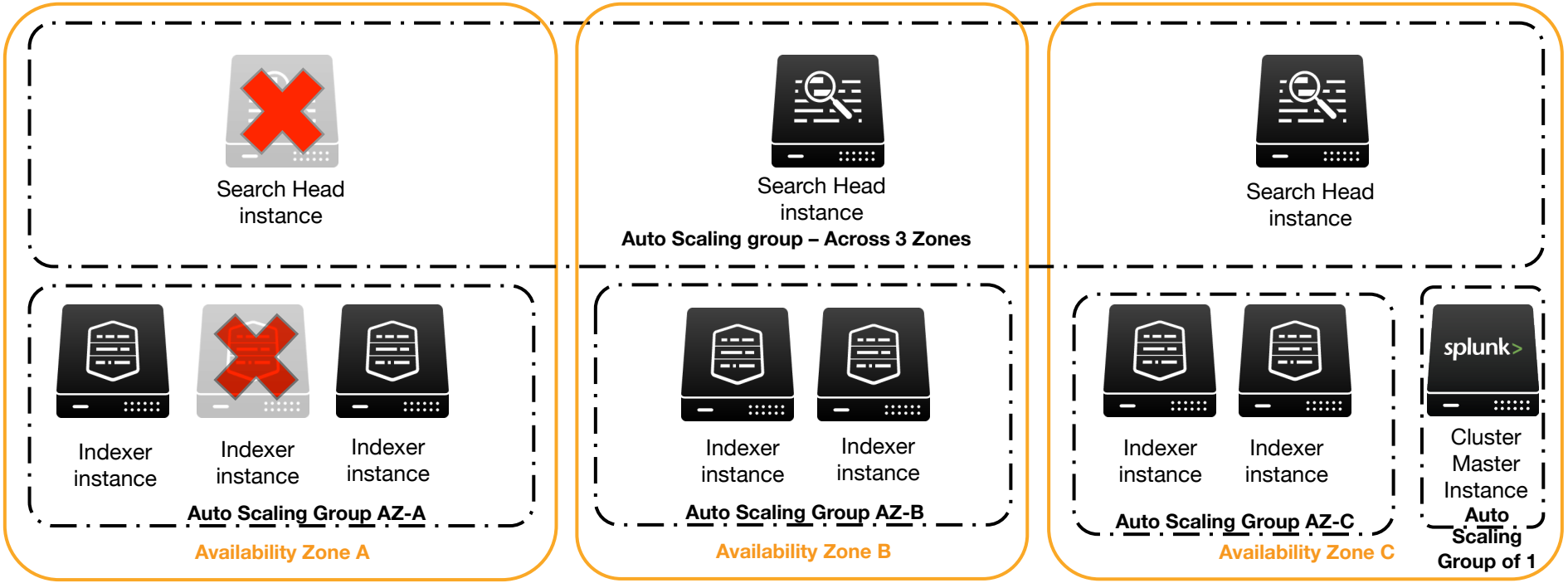
Splunk Search Head Clustering with Auto Scaling

- Auto-election of captain within the Search Head Cluster
- Auto Scaling Policy spans across multiple Availability Zones

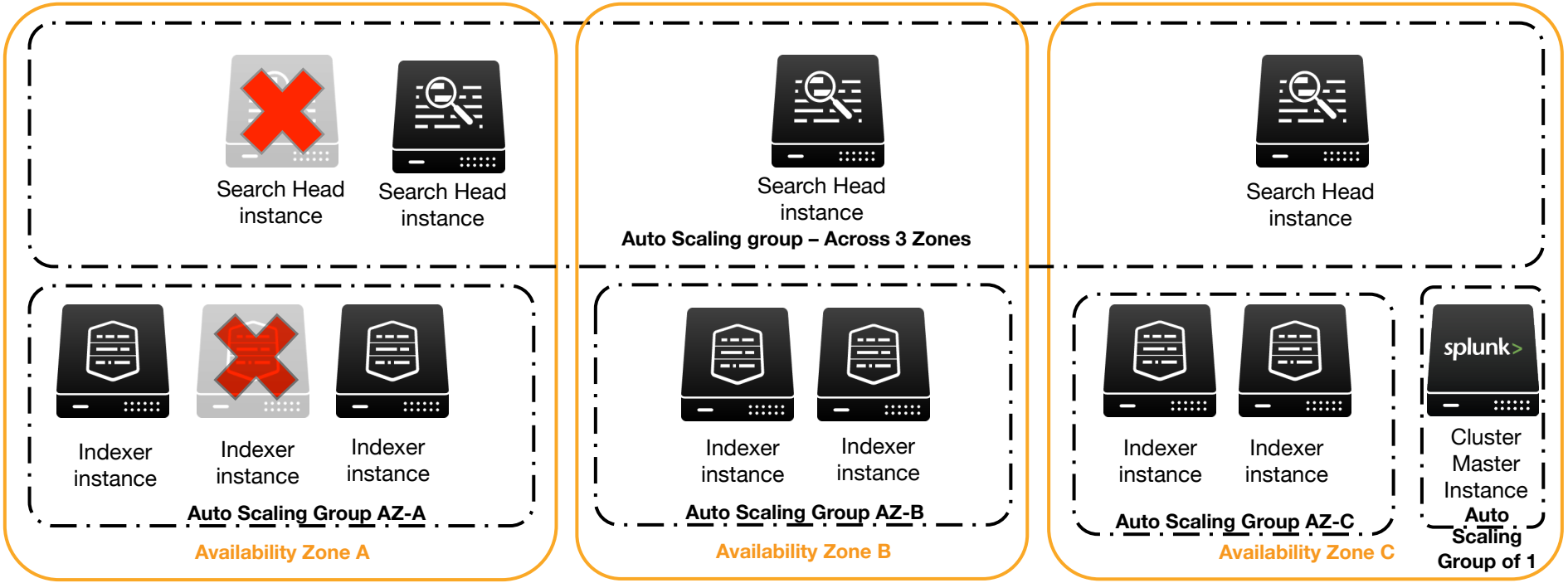
Architecture Diagram (Splunk + AWS)



Architecture Diagram (Splunk + AWS)



Architecture Diagram (Splunk + AWS)



Splunk + AWS Features = FTW

- Self Healing Splunk Infrastructure
- Splunk Clustering provides data availability and replication
- AWS Auto Scaling can automatically replace failed Splunk instances

Splunk + AWS Auto Scaling Considerations

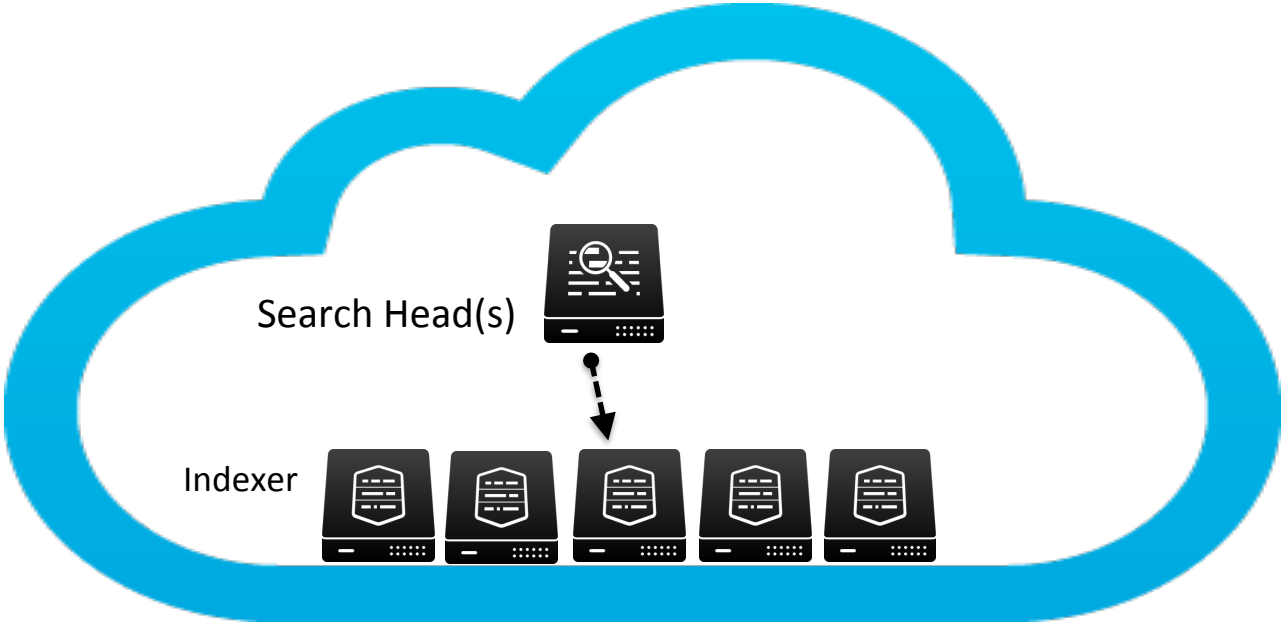
- Auto Scaling Group of 1 for Splunk Cluster Master
  - Splunk Cluster Master is a stateless server
- Use DNS name instead of IP address for Splunk Cluster Master URI
- Bootstrap EC2 instances to automatically join Splunk Indexer and Search Head Clusters

Provisioning with CloudFormation

Why CloudFormation?

- Fast, automated, consistent Splunk deployments on AWS
- Incorporates Splunk **best practices** for operations and administration
- **Abstracts away** details of configuring distributed Splunk
- **Extensible** and **customizable** templates to fit custom needs
- It's free

CloudFormation Deployment



Splunk + AWS + CloudFormation
Ready in minutes…

Splunk App for AWS Demo

Splunk App for AWS

- Splunk Add-on for AWS
  - Data Collection for Cloudtrail, Cloudwatch, Config & Config rules, EC2, ELB, EBS, S3, Billing, Inspector, RDS, etc.
  - Splunk analyzes data from various AWS sources
- Splunk App for AWS
  - Analytics and dashboards of AWS ecosystem

"Customers love having the agility of AWS with the end-to-end visibility of Splunk."

- Andy Jassy

CEO, AWS

Splunk AWS App Demo

Questions?

Contact

Simeon Yep
syep@splunk.com

Nate Kwong
nkwong@splunk.com

Bill Bartlett
bbartlett@splunk.com

References

- Splunk App for AWS: http://apps.splunk.com/app/1274/
- Hunk App for AWS ELB: http://apps.splunk.com/app/1731/
- Technical Brief: http://www.splunk.com/content/dam/splunk2/pdfs/technical-briefs/deploying-splunk-enterprise-on-amazon-web-services-technical-brief.pdf
- How Autodesk Leverages Splunk on AWS - re:Invent session: https://www.youtube.com/watch?v=ofYgkqK-fLE

References

- Blogs:
  - http://blogs.splunk.com/2012/03/07/splunk-and-aws-sizing-revisited/
  - http://blogs.splunk.com/2013/06/06/splunkit-v2-0-2-results-ec2-storage-comparisons/
  - http://blogs.splunk.com/2013/07/31/whats-going-on-with-aws-and-splunk/
  - http://blogs.splunk.com/2014/05/20/deploy-your-own-splunk-cluster-on-aws-in-minutes/
- AMIs
  - Splunk: https://aws.amazon.com/marketplace/pp/B00GIZITUO?sr=0-4
  - Hunk: https://aws.amazon.com/marketplace/pp/B00GIZK2QI?sr=0-2

THANK YOU

splunk>